

Standard flow liquid chromatography for shotgun proteomics in bioenergy research

Susana M. González Fernández-Niño¹, A. Michelle Smith-Moritz¹, Leanne Jade G. Chan¹, Paul D. Adams^{1,2}, Joshua L. Heazlewood^{1,3} and Christopher J. Petzold^{1*}

¹ Joint BioEnergy Institute and Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

² Department of Bioengineering, University of California Berkeley, Berkeley, CA, USA

³ Australian Research Council Centre of Excellence in Plant Cell Walls, School of Botany, The University of Melbourne, Melbourne, VIC, Australia

Edited by:

Robert Henry, The University of Queensland, Australia

Reviewed by:

Qaisar Mahmood, COMSATS Institute of Information Technology, Pakistan

Yu-Shen Cheng, University of California Davis, USA

*Correspondence:

Christopher J. Petzold, Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, 5885 Hollis Street, 4th Floor, Emeryville, CA 94608, USA
e-mail: cjpetzold@lbl.gov

Over the past 10 years, the bioenergy field has realized significant achievements that have encouraged many follow on efforts centered on biosynthetic production of fuel-like compounds. Key to the success of these efforts has been transformational developments in feedstock characterization and metabolic engineering of biofuel-producing microbes. Lagging far behind these advancements are analytical methods to characterize and quantify systems of interest to the bioenergy field. In particular, the utilization of proteomics, while valuable for identifying novel enzymes and diagnosing problems associated with biofuel-producing microbes, is limited by a lack of robustness and limited throughput. Nano-flow liquid chromatography coupled to high-mass accuracy, high-resolution mass spectrometers has become the dominant approach for the analysis of complex proteomic samples, yet such assays still require dedicated experts for data acquisition, analysis, and instrument upkeep. The recent adoption of standard flow chromatography (ca. 0.5 mL/min) for targeted proteomics has highlighted the robust nature and increased throughput of this approach for sample analysis. Consequently, we assessed the applicability of standard flow liquid chromatography for shotgun proteomics using samples from *Escherichia coli* and *Arabidopsis thaliana*, organisms commonly used as model systems for lignocellulosic biofuels research. Employing 120 min gradients with standard flow chromatography, we were able to routinely identify nearly 800 proteins from *E. coli* samples; while for samples from *Arabidopsis*, over 1,000 proteins could be reliably identified. An examination of identified peptides indicated that the method was suitable for reproducible applications in shotgun proteomics. Standard flow liquid chromatography for shotgun proteomics provides a robust approach for the analysis of complex samples. To the best of our knowledge, this study represents the first attempt to validate the standard flow approach for shotgun proteomics.

Keywords: proteomics, standard flow chromatography, biofuels, mass spectrometry

INTRODUCTION

Advances in biofuels research focusing on feedstock characterization and engineering (Persil-Cetinkol et al., 2012; DeMartini et al., 2013; Shen et al., 2013; Eudes et al., 2014) as well as the genetic manipulation of microbes (Alper et al., 2006; Tyo et al., 2007; Keasling, 2008; Lee et al., 2008) have progressed significantly in the last few years. Unfortunately, analytical capabilities required to efficiently monitor and assess these changes are lagging. Furthermore, many modern bioanalytical techniques are focused toward medical and health related research, which have significantly different priorities and requirements for success. Most biotechnology research challenges are not constrained by sensitivity or resolution of an assay, rather they depend on accurate identification and quantitation of target molecules for a large number of samples. Consequently, an important component for biotechnological research is sample throughput supported by a robust analytical platform. Recent advances in proteomics and metabolomics have

focused on liquid chromatography-mass spectrometry (LC-MS) methods by increasing their sensitivity to aid discovery-based research efforts. This is most evident with the development of nano-LC couple to high-resolution mass spectrometers; yet, this technology is yet to mature into a robust platform capable of consistently analyzing hundreds of samples per week. Consequently, alternate technologies capable of answering the questions necessary for biotechnology progress are needed. Recently, we published a high throughput targeted proteomic toolkit based on standard flow chromatography coupled to mass spectrometry (MS) to help address these issues for *Escherichia coli*; however, a significant amount of methods development was necessary. This work prompted us to assess the utility of standard flow liquid chromatography (LC) for shotgun proteomic methods related to biotechnology.

The discipline of proteomics has been dominated by nano-flow LC coupled to MS since its early development over 20 years ago

(Emmett and Caprioli, 1994; Gatlin et al., 1998). The adoption of nano-flow LC for protein identification was driven by the substantial increases in sensitivity and detection capabilities of nano-flow (ca. 500 nL/min) over capillary (ca. 50 μ L/min) and standard flow (ca. 0.5 mL/min) chromatography. Typically, shotgun proteomic studies utilize nano-flow chromatography methods due to limited amounts of sample and to obtain optimal ionization efficiency for sensitive detection of peptides but at the cost of ease of use and system robustness (Gapeev et al., 2009). These issues are particularly problematic for biotechnology research that depends heavily on high sample throughput. Recently, ultra-high performance LC coupled to triple quadrupole mass spectrometers has been shown to yield comparable sensitivity and better analytical metrics (coefficient of variation, dynamic range) than nano-flow LC for MRM-based analysis of biomarker proteins when the amount of sample is adjusted to the column size (Percy et al., 2012). Additionally, the well-established robust operation of standard flow chromatography makes this instrumentation attractive for applications that rely on high sample throughput, consistent results, and less system downtime (Swartz, 2005). For applications where sample abundance is not limited, one of the greatest concerns with ultra-high pressure liquid chromatography-mass spectrometry (UHPLC-MS) workflows is the loss of sensitivity, relative to nano-LC-MS workflows, leading to datasets of insufficient depth to answer questions of interest.

The adoption of a UHPLC-MS workflow for sample delivery requires both efficient ionization and instrument speed to handle both the sample delivery rate and reduced elution times for peptides. The past decade has seen significant developments and advances in instrumentation associated with proteomics-based MS. Advances in reversed phase C18 chromatography columns yield greater separation efficiency and more stable retention times. The current generation of instrumentation is faster, more sensitive, and is better able to deal with the dynamic range inherent in biological samples. Moreover, the development and adoption of off-axis nebulizers for sample delivery when using electrospray ionization significantly reduces contamination of capillaries and skimmers (Banerjee and Mazumdar, 2012). Collectively, these improvements have enabled the development of a UHPLC-MS workflow for MRM-based analysis of biomarker proteins (Percy et al., 2012). Consequently, we were interested in assessing the capacity of this workflow in shotgun proteomic experiments. The work described here details the results of an analysis of a prokaryote (*E. coli*) and an eukaryote (*Arabidopsis thaliana*) whole cell proteomes on an Agilent UHPLC-QTOFMS system, but would be generally applicable for any current generation of tandem mass spectrometer being utilized for shotgun proteomics.

MATERIALS AND METHODS

PROTEIN EXTRACTION

Protein was extracted using standard techniques with analytical reagents where suitable. For *E. coli* DH5 α samples, cell lysis and protein precipitation was accomplished using a chloroform/methanol precipitation. A 100 μ L aliquot of cells was transferred to a 1.7 mL tube, followed by the addition of 400 μ L of methanol, 100 μ L of chloroform, and 300 μ L of water, with mixing by vortex after each addition. Following centrifugation at

21,000 $\times g$ for phase separation, the methanol and water layer was removed and 300 μ L of methanol was added. The tube was briefly vortexed to dislodge the protein pellet, then centrifuged at 21,000 $\times g$ for 2 min. The chloroform and methanol layer was removed and the protein pellet was dried for 5 min in a vacuum concentrator. The protein pellet was re-suspended in 100 mM (NH₄)HCO₃ with 20% methanol, reduced with 5 mM TCEP [Tris(2-carboxyethyl)phosphine hydrochloride] for 30 min at room temperature, treated with 10 mM iodoacetamide (IAA) for 30 min in the dark at room temperature, and digested with trypsin (1:50 w/w) overnight at 37°C. Aliquots of 40 μ g were taken for analysis by LC-MS/MS. For *A. thaliana* (L.) Heynh. (ecotype Landsberg erecta), protein was extracted from a previously described heterotrophic cell culture (Ito et al., 2011). A total of 1 g plant material (fresh weight) was used for the isolation of total protein. The plant material was harvested and frozen with liquid nitrogen in an Eppendorf tube with two small steel balls. The protein extraction was performed by the addition of 0.4 mL of fresh disruption buffer [125 mM Tris-HCl, 7% (w/v) SDS, and 10% β -mercaptoethanol], followed by vortex for 10 min at room temperature. The samples were centrifuged at 10,000 $\times g$ for 5 min at 4°C and the supernatant separated into two 2 mL tubes. Samples were further extracted in 800 μ L methanol and mixed, then 200 μ L chloroform added and mixed, and finally 500 μ L of ddH₂O added and vortexed (30 sec each time). Samples were centrifuged for 5 min at 10,000 $\times g$ at 4°C, the aqueous phase removed, and 500 μ L of methanol added. Samples were vortexed for 30 s and centrifuged for 10 min at 9,000 $\times g$ at 4°C. The supernatant was discarded and the pellet air-dried. The dried pellet was suspended in 200 μ L of re-suspension buffer [3M urea, 50 mM (NH₄)HCO₃, and 5 mM dithiothreitol, pH 8], and incubated with IAA at a final concentration 10 mM for 30 min in the dark. Prior to analysis by MS, 40 μ g of extracted protein was digested with trypsin (1:10 w/w) overnight at 37°C. Peptides were desalted using C18 Micro SpinColumns (Harvard Apparatus) as previously outlined (Parsons et al., 2013). Eluted peptides were re-suspended in 2% acetonitrile, 0.1% formic acid prior to analysis by LC-MS/MS.

STANDARD FLOW MASS SPECTROMETRY

All samples were analyzed on an Agilent 6550 iFunnel Q-TOF mass spectrometer (Agilent Technologies) coupled to an Agilent 1290 UHPLC system. Peptide samples were loaded onto a Sigma-Aldrich Ascentis Peptides ES-C18 column (2.1 mm \times 100 mm, 2.7 μ m particle size, operated at 60°C) via an Infinity Autosampler (Agilent Technologies) with Buffer A (2% acetonitrile, 0.1% formic acid) flowing at 0.400 mL/min. Peptides were eluted into the mass spectrometer via a gradient with initial starting condition of 5% buffer B (98% acetonitrile, 0.1% formic acid). For analysis of all samples, buffer B was increased to 35% over 120 min. Buffer B was then increased to 50% over 5 min, then up to 90% over 1 min, and held for 7 min at a flow rate of 0.6 mL/min, followed by a ramp back down to 5% B over 1 min where it was held for 6 min to re-equilibrate the column to original conditions. Peptides were introduced to the mass spectrometer from the LC by using a Jet Stream source (Agilent Technologies) operating in positive-ion mode (3,500 V). Source parameters employed gas temp (250°C), drying gas (14 L/min), nebulizer (35 psig), sheath

gas temp (250°C), sheath gas flow (11 L/min), VCap (3,500 V), fragmentor (180 V), OCT 1 RF Vpp (750 V). The data were acquired with Agilent MassHunter Workstation Software, LC/MS Data Acquisition B.05.00 (Build 5.0.5042.2) operating in Auto MS/MS mode whereby the 20 most intense ions (charge states, 2–5) within 300–1,400 m/z mass range above a threshold of 1,500 counts were selected for MS/MS analysis. MS/MS spectra (100–1,700 m/z) were collected with the quadrupole set to “Medium” resolution and were acquired until 45,000 total counts were collected or for a maximum accumulation time of 333 ms. Former parent ions were excluded for 0.1 min following MS/MS acquisition.

LC-MS/MS DATA ANALYSIS AND INTEGRATION

The acquired data were exported as .mgf files using the Export as MGF function of the MassHunter Workstation Software, Qualitative Analysis (Version B.05.00 Build 5.0.519.13 Service Pack 1, Agilent Technologies) using the following settings: peak Filters (MS/MS), the absolute height (≥ 20 counts), relative height ($\geq 0.100\%$ of largest peak), maximum number of peaks (300) by height; for charge state (MS/MS), the peak spacing tolerance (0.0025 m/z plus 7.0 ppm), isotope model (peptides), charge state limit assigned to (5) maximum. Resultant data files were interrogated with the Mascot search engine version 2.3.02 (Matrix Science) with a peptide tolerance of ± 50 ppm and MS/MS tolerance of ± 0.1 Da; fixed modifications Carbamidomethyl (C); variable modifications Oxidation (M); up to one missed cleavage for trypsin; peptide charge 2+, 3+, and 4+; and the instrument type was set to ESI-QUAD-TOF. Searches were performed against either an *E. coli* (strain K12) dataset obtained from UniProt (Magrane and Consortium, 2011) or the latest release of the *A. thaliana* dataset comprising TAIR10 obtained from The Arabidopsis Information Resource (Lamesch et al., 2012). Both databases incorporated proteins comprising the common Repository of Adventitious Proteins (cRAP v2012.01.01 from The Global Proteome Machine). The *E. coli* database comprised 4,429 sequences (1,398,775 residues) while the *Arabidopsis* database comprised 35,508 sequences (14,522,421 residues). Protein and peptide matches identified after interrogation of MS/MS data by Mascot were filtered and validated using Scaffold v4.3.0 (Proteome Software Inc.). Peptide identifications were accepted if they could be established at $>95.0\%$ probability by the Peptide Prophet algorithm (Keller et al., 2002) with Scaffold delta-mass correction. Protein identifications were accepted if they could be established at $>95.0\%$ probability and contained at least 1 identified peptide (at 95% and greater). Protein probabilities were assigned by the Protein Prophet algorithm (Nesvizhskii et al., 2003). This resulted in false discovery rates of 0.9 (*E. coli*) and 0.3% (*Arabidopsis*) for protein and 0.37 (*E. coli*) and 1.66% (*Arabidopsis*) for peptides. Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony.

STATISTICAL DATA ANALYSIS

Statistical analysis was performed using the statistical toolbox in MATLAB v2009b (Mathworks). The following analysis was performed on data from both *Arabidopsis* and *E. coli*. After data

filtering using Scaffold v4.3.0 (outlined above), peptide data were exported as .csv files and only peptides derived from the same protein identified across all replicates were considered. The Mascot ions score and total ion current values were only used for the best matching peptide (based on ion score) in each replicate. The total spectrum count for each peptide from each replicate was calculated manually. The coefficient of variation (CV: SD/mean) of total spectrum counts, Mascot ions score, and total ion current for each of the common peptides across the replicates was calculated to determine the variation across the technical replicates. The histograms of the CV were plotted to determine distribution and general extreme value fit algorithm was used to determine fit and mean of the distribution. For principal component analysis (PCA), the total spectrum count, Mascot ions score, and total ion current for all the common peptides of each technical replicate were mean centered. PCA was done by eigenvalue decomposition of the data covariance, resulting in a set of linearly uncorrelated variables. The principal component scores were then plotted to identify groupings.

RESULTS AND DISCUSSION

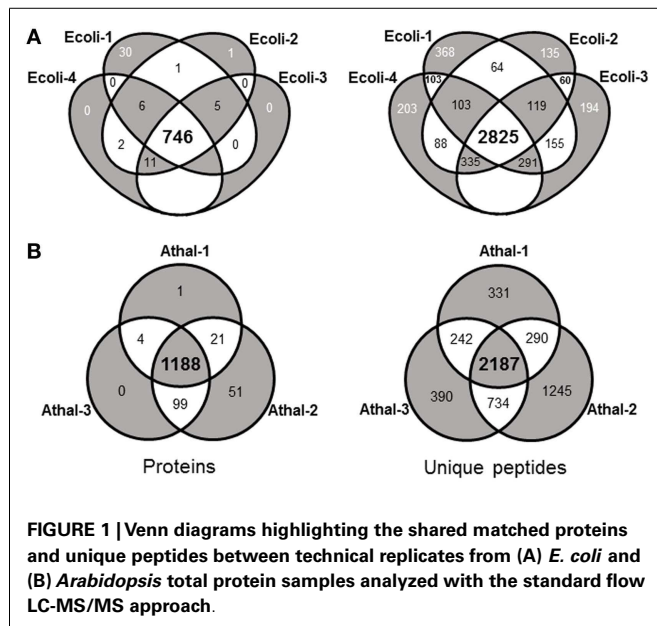
The initial setup and tuning of the standard flow LC-MS/MS parameters was undertaken using 20 fmol aliquots of trypsin digested bovine serum albumin (BSA). The system settings were deemed adequate once identification of BSA was comparable (i.e., unique peptides, total peptides, and coverage) to that achieved using previously benchmarked nano-flow LC-MS/MS approaches on a variety of instruments over the past 5 years.

APPLICATION OF STANDARD FLOW LC-MS/MS WITH PROKARYOTIC SAMPLES

Initial experiments were performed on *E. coli*, a well-characterized organism with minimal proteomic complexity. The *E. coli* genome of the widely utilized laboratory strain K-12 was completed nearly 20 years ago (Blattner et al., 1997). The genome has undergone multiple revisions and is estimated to codes for over 4,300 proteins (Magrane and Consortium, 2011). Shake flasks of *E. coli* DH5 α cultures were grown aerobically at 37°C on Luria Broth (LB) medium supplemented with 1% glucose. Total protein was extracted and digested overnight with trypsin at 37°C. The analysis procedure for *E. coli* samples was developed to take advantage of the speed and robustness of a standard flow analysis and as such, a method was developed that incorporated a 120 min gradient. A total of four biological replicates each equivalent to 40 μ g of peptide were analyzed by a standard flow UHPLC-QTOFMS operating with typical shotgun proteomics data acquisition parameters. The samples yielded on average $31,938 \pm 989$ (SE) MS/MS spectra (Table 1). Each of these MS/MS datasets was used to query the *E. coli* (K12) protein database from UniProt (Magrane and Consortium, 2011) using the Mascot search engine (Matrix Science) with resultant protein matches filtered using Scaffold (Proteome Software). In total, 802 proteins were identified from the four replicates with an average of 786 ± 4.5 (SE) per sample (Table 1 and Table S1 in Supplementary Material). This represents about 20% of the protein coding capacity of *E. coli*. On average, $14,639 \pm 484$ (SE) MS/MS spectra were successfully matched to a peptide, corresponding to around 46% of the total queries for

Table 1 | Values obtained from *E. coli* (Ec) and *Arabidopsis* (At) samples analyzed using the standard flow technique.

	Ec 1	Ec 2	Ec 3	Ec 4	At 1	At 2	At 3
Queries	31,452	29,348	33,638	33,315	26,135	37,388	28,203
Spectra	14,677	13,265	15,439	15,173	6,038	9,712	6,846
Spectra matched (%)	47	45	46	46	23	26	24
Unique peptides	4,028	3,729	4,154	4,123	3,050	6,643	3,553
Total proteins	788	772	791	791	1,214	1,359	1,291

**FIGURE 1 | Venn diagrams highlighting the shared matched proteins and unique peptides between technical replicates from (A) *E. coli* and (B) *Arabidopsis* total protein samples analyzed with the standard flow LC-MS/MS approach.**

each sample (Table 1). This would be regarded as a high conversion rate indicating that the acquired MS/MS was of sufficient quality to confidently assign nearly 50% of the queries using the approach. The consistency of the standard flow approach was highlighted by the total number of proteins assigned in each sample. A total of 768 proteins were identified in three out of four replicates, while 746 proteins (93%) were identified collectively in all four samples. This compares well to the total number of proteins identified in all samples (802). Few proteins were identified uniquely in a single replicate except for replicate Ecoli-1, where 30 unique proteins were identified (Figure 1A). Consequently, we sought to examine whether this setup was adequately dealing with the sample complexity given the flow rate and an approximate peptide elution time of around 6 s (Figure S1 in Supplementary Material). Nearly 75% of the matched MS/MS spectra were redundant, with an average of 4,009 unique peptides per sample (Table 1). Only 2,825 unique spectra (54%) were shared between all four samples; however, the new unique peptides (2,393) accounted for only 56 new proteins (Figure 1A). The approach consistently identifies the major proteins in a sample and a majority of unique spectra assigned between replicates are derived from previously identified proteins. This indicates that the QTOFMS has the capacity to handle the standard flow UHPLC setup in shotgun mode even at these higher flow rates and reduced peptide elution times.

APPLICATION OF STANDARD FLOW LC-MS/MS TO A COMPLEX EUKARYOTIC SAMPLE

Next, we were interested in assessing the suitability of the standard flow proteomics platform on a complex eukaryotic proteome. The genome of the reference plant *Arabidopsis* was completed over a decade ago (*Arabidopsis* Genome Initiative, 2000) and now represents one of the most highly curated and annotated genomes in biology. As a consequence, *Arabidopsis* is being utilized as a proving ground for plant synthetic biology approaches, many of which have focused on biomass manipulation for efficient biofuels production (Eudes et al., 2012, 2015). The most recent proteome release (TAIR10) from The *Arabidopsis* Information Resource (Lamesch et al., 2012) comprises over 27,000 loci and over 35,000 distinct protein products. Protein extractions were performed from a 7-day old *Arabidopsis* cell cultures that have previously been extensively employed for proteomic assays (Parsons et al., 2012). A total protein sample was analyzed in triplicate by UHPLC-QTOFMS using the standard flow proteomics platform. We analyzed the equivalent of ca. 40 μ g of digested total protein over a 120 min gradient with MS conditions identical to those used with the *E. coli* samples. An average of $30,575 \pm 3,458$ (SE) MS/MS spectra were collected over the 2-hour run from the three replicates (Table 1). These numbers are similar to those obtained using the *E. coli* samples and may reflect the upper capacity of the method given the increased peptide complexity that would be expected from a eukaryotic sample.

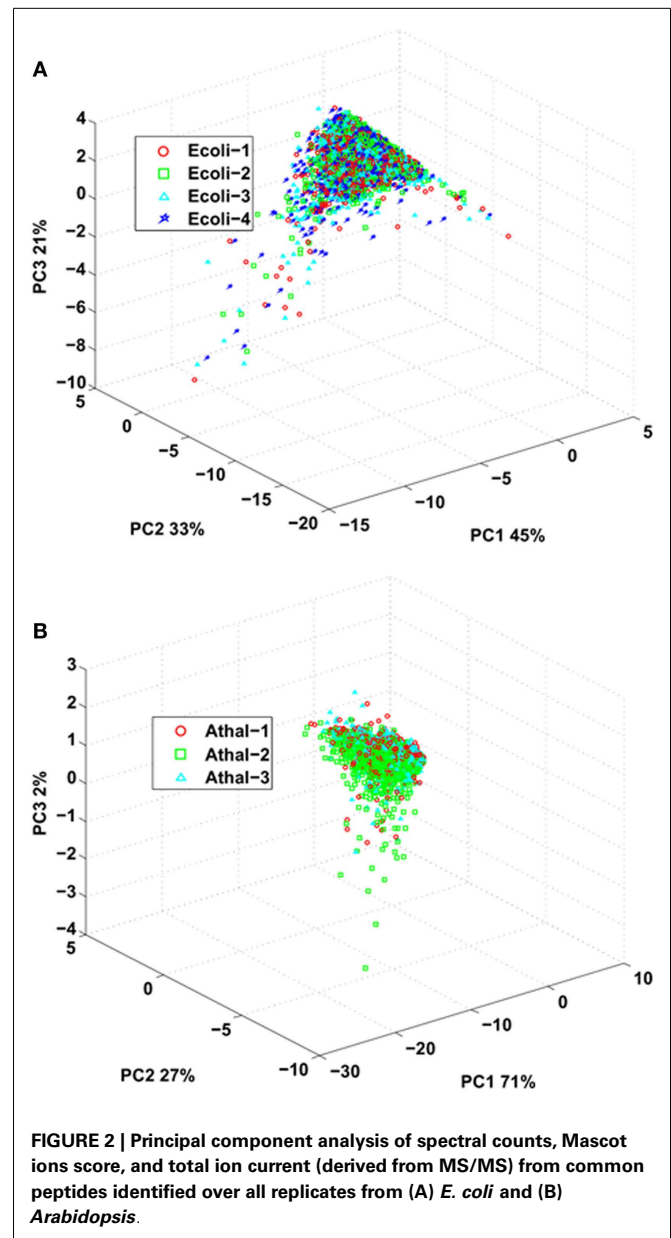
These *Arabidopsis* datasets were each used to interrogate the most recent *Arabidopsis* protein dataset using the Mascot search engine and matches filtered and integrated using Scaffold. This resulted in an average of $7,532 \pm 1,115$ (SE) matched spectra from the three replicates and corresponds to about 25% of the total spectra obtained (Table 1). This corresponds to about half the conversion rate observed for the *E. coli* samples and likely reflects the quality of the MS/MS spectra given the overall ion intensities are likely on average considerably lower due to the increased size of the proteome. The average number of proteins identified over the three samples was $1,288 \pm 73$ (SE) with a total of 1,364 unique proteins identified in all three replicates (Table S2 in Supplementary Material). The number of proteins consistently identified by all three replicates was 1,188 proteins or 87% of the total number (Figure 1B). The minor variation in identifications between these technical replicates demonstrates the reproducibility of the standard flow approach for shotgun proteomics as the method was capable of consistently identify the same proteins in each of the replicates. In an attempt to understand whether the approach is adequately dealing with the increased complexity of this eukaryotic sample, we examined the proportion of unique

matched peptides identified in each replicate. The 1,364 unique proteins identified in these samples were matched using 5,419 unique peptides, while the 1,188 proteins identified in all three replicates only required 2,187 (40.4% of the total) unique peptides (**Figure 1B**). These results indicate that again, as found with the *E. coli* samples, the majority of peptides exclusive to a given replicate are derived from proteins that have already yielded a high scoring peptide match. It is conceivable that the complexity of the eukaryotic sample contributes to many more co-eluting peptides during the 120 min gradient than the *E. coli* sample, resulting in unique peptides selected for MS/MS by the data-dependent acquisition method in each replicate (**Figure 1B**). Although similar results with regard to unique peptides were also observed with the *E. coli* samples, on average, an *E. coli* protein was identified by 6.8 unique peptides (768 proteins from 5,218 unique peptides) compared to the *Arabidopsis* proteins at 4.0 unique peptides each (1,364 proteins from 5,419 unique peptides). Nearly twice as many proteins were identified in the *Arabidopsis* samples (1,364 proteins compared to 768 proteins), which would be expected given the differences in coding capacity between these species. However, considerably more identified proteins would have been expected given proteome sizes, indicating that the standard flow approach has a more limited capacity to deal with the eukaryotic sample due to issues, such as increased sample complexity at any given point in time, lower overall ion intensity, and dynamic range limitations (as discussed above) of this QTOFMS system.

ASSESSING THE PERFORMANCE OF STANDARD FLOW LC-MS/MS

Last, we investigated the reproducibility of the standard flow UHPLC-QTOFMS by comparing parameters of the identified peptides between the replicates for both *E. coli* and *Arabidopsis* samples. The peptide ions score obtained from Mascot after data interrogation can be indicative of the quality of the fragmentation spectra (Perkins et al., 1999), total ion current obtained from the MS/MS spectra can provide information about the intensity of an eluted peptide (Asara et al., 2008) while total spectral count can provide an indication of peptide intensity and spectral complexity (Lundgren et al., 2010). These values can be used as a proxy to assess sample limitations and reproducibility by the LC-MS system. PCA was employed to ascertain whether there were any differences between ion scores, total ion current, and spectral count for peptides identified in all the replicates for either *E. coli* or *Arabidopsis* samples (**Figure 2**). The analysis demonstrated that none of the replicates for either *E. coli* or *Arabidopsis* could be separated by principal component scores for these attributes, indicating that majority of the differences for either data set can be largely attributed to biological heterogeneity of the sample (**Figure 2**).

The similarities in ion score, total ion current, and spectral counts can also be observed when analyzing the distributions of CV for the each identified peptides across the replicates (Figure S2 in Supplementary Material). The variations in Mascot ion score and total ion current (derived from MS/MS spectra) for the peptides identified over the replicates were similar for both *E. coli* and *Arabidopsis*. This indicates that from sample to sample, identical ions performed similarly with regard to the intensity of matched



peptides (total ion current) as measured by the mass spectrometer. This is further supported by the small variation in Mascot ion scores with ion intensity having a relationship to the quality of the MS/MS spectra and subsequent spectral matching procedures. The variation in total spectrum count was more pronounced between the identified peptides of *E. coli* (0.67) and *Arabidopsis* (0.32), with samples of lower complexity (i.e., *E. coli*) having a larger variation in spectral counts for a peptide across replicates. This could be due to the higher repeat sampling rate that likely occurred during analysis of *E. coli* samples due to the reduced number of distinct ions/peptides in the sample. This conclusion is supported when looking at the average number of spectral counts for a given identified peptide; for *E. coli*, it was 3.74 spectra per peptide while for *Arabidopsis* it was 2.70. The lower average number of spectral

counts for a peptide from *Arabidopsis* is likely indicative of the increased sample complexity. Collectively, these data indicate that between the replicates, the quality and intensities of peptides identified across the replicates was similar, indicating that the standard flow approach was not significantly impeding the performance of the mass spectrometer to acquire tandem mass spectra.

CONCLUSION

The recent progress of bioenergy research has relied heavily on transformational developments in feedstock characterization and metabolic engineering. Yet, omics methods to characterize and quantify systems of interest have mainly been adapted from the health and clinical fields that have very different research needs (i.e., high sensitivity, deep proteome coverage). We report the application of a standard flow LC-MS/MS approach that is suitable for large numbers of shotgun proteomic experiments where sample abundance is not limiting. The setup is capable of undertaking a rapid analysis of low complexity samples as well as handling highly complex samples by employing extended analysis times. Although its application requires instrumentation with the ability to deal with increased flow rates and shorter peptide elution times, it is apparent that the current generation of tandem mass spectrometers is capable of handling these parameters. While traditional nano-LC-MS/MS approaches are likely to continue to dominate shotgun analyses as they produce a greater number of protein identifications and have the ability to deal with increased sample complexity and dynamic range, the robust nature and simplicity of standard flow coupled to MS makes this approach an attractive alternative for applications where sample throughput and reproducibility are important factors.

AUTHOR CONTRIBUTIONS

JH and CP conceived and advised in all aspects of the study. PA supervised all aspects of the study. SG and LC performed experiments. AS analyzed the data. JH and CP interpreted the data and wrote the manuscript. All authors discussed and commented on the manuscript.

ACKNOWLEDGMENTS

This work was part of the DOE Joint BioEnergy Institute (<http://www.jbei.org>) supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at <http://journal.frontiersin.org/article/10.3389/fbioe.2015.00044/abstract>

REFERENCES

- Alper, H., Moxley, J., Nevoigt, E., Fink, G. R., and Stephanopoulos, G. (2006). Engineering yeast transcription machinery for improved ethanol tolerance and production. *Science* 314, 1565–1568. doi:10.1126/science.1131969
- Arabidopsis Genome Initiative. (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815. doi:10.1038/35048692
- Asara, J. M., Christofk, H. R., Freemark, L. M., and Cantley, L. C. (2008). A label-free quantification method by MS/MS TIC compared to SILAC and spectral counting in a proteomics screen. *Proteomics* 8, 994–999. doi:10.1002/pmic.200700426
- Banerjee, S., and Mazumdar, S. (2012). Electrospray ionization mass spectrometry: a technique to access the information beyond the molecular weight of the analyte. *Int. J. Anal. Chem.* 2012, 282574. doi:10.1155/2012/282574
- Blattner, F. R., Plunkett, G. III, Bloch, C. A., Perna, N. T., Burland, V., Riley, M., et al. (1997). The complete genome sequence of *Escherichia coli* K-12. *Science* 277, 1453–1462. doi:10.1126/science.277.5331.1453
- DeMartini, J. D., Pattathil, S., Miller, J. S., Li, H. J., Hahn, M. G., and Wyman, C. E. (2013). Investigating plant cell wall components that affect biomass recalcitrance in poplar and switchgrass. *Energy Environ. Sci.* 6, 898–909. doi:10.1039/c3ee23801f
- Emmett, M. R., and Caprioli, R. M. (1994). Micro-electrospray mass spectrometry: ultra-high-sensitivity analysis of peptides and proteins. *J. Am. Soc. Mass Spectrom.* 5, 605–613. doi:10.1016/1044-0305(94)85001-1
- Eudes, A., George, A., Mukerjee, P., Kim, J. S., Pollet, B., Benke, P. I., et al. (2012). Biosynthesis and incorporation of side-chain-truncated lignin monomers to reduce lignin polymerization and enhance saccharification. *Plant Biotechnol. J.* 10, 609–620. doi:10.1111/j.1467-7652.2012.00692.x
- Eudes, A., Liang, Y., Mitra, P., and Loque, D. (2014). Lignin bioengineering. *Curr. Opin. Biotechnol.* 26, 189–198. doi:10.1016/j.copbio.2014.01.002
- Eudes, A., Sathitsuksanoh, N., Baidoo, E. E. K., George, A., Liang, Y., Yang, F., et al. (2015). Expression of a bacterial 3-dehydroshikimate dehydratase reduces lignin content and improves biomass saccharification efficiency. *Plant Biotechnol. J.* doi:10.1111/pbi.12310
- Gapeev, A., Berton, A., and Fabris, D. (2009). Current-controlled nanospray ionization mass spectrometry. *J. Am. Soc. Mass Spectrom.* 20, 1334–1341. doi:10.1016/j.jasms.2009.03.007
- Gatlin, C. L., Kleemann, G. R., Hays, L. G., Link, A. J., and Yates, J. R. III (1998). Protein identification at the low femtomole level from silver-stained gels using a new fritless electrospray interface for liquid chromatography-microspray and nanospray mass spectrometry. *Anal. Biochem.* 263, 93–101. doi:10.1006/abio.1998.2809
- Ito, J., Bath, T. S., Petzold, C. J., Redding-Johanson, A. M., Mukhopadhyay, A., Verboom, R., et al. (2011). Analysis of the *Arabidopsis* cytosolic proteome highlights subcellular partitioning of central plant metabolism. *J. Proteome Res.* 10, 1571–1582. doi:10.1021/pr1009433
- Keasling, J. D. (2008). Synthetic biology for synthetic chemistry. *ACS Chem. Biol.* 3, 64–76. doi:10.1021/cb7002434
- Keller, A., Nesvizhskii, A. I., Kolker, E., and Aebersold, R. (2002). Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* 74, 5383–5392. doi:10.1021/ac025747h
- Lamesch, P., Berardini, T. Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., et al. (2012). The *Arabidopsis* information resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.* 40, D1202–D1210. doi:10.1093/nar/gkr1090
- Lee, S. K., Chou, H., Ham, T. S., Lee, T. S., and Keasling, J. D. (2008). Metabolic engineering of microorganisms for biofuels production: from bugs to synthetic biology to fuels. *Curr. Opin. Biotechnol.* 19, 556–563. doi:10.1016/j.copbio.2008.10.014
- Lundgren, D. H., Hwang, S. I., Wu, L. F., and Han, D. K. (2010). Role of spectral counting in quantitative proteomics. *Expert. Rev. Proteomics* 7, 39–53. doi:10.1586/epr.09.69
- Magrane, M., and Consortium, U. (2011). UniProt knowledgebase: a hub of integrated protein data. *Database(Oxford)* 2011, bar009. doi:10.1093/database/bar009
- Nesvizhskii, A. I., Keller, A., Kolker, E., and Aebersold, R. (2003). A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* 75, 4646–4658. doi:10.1021/ac0341261
- Parsons, H. T., Christiansen, K., Knierim, B., Carroll, A., Ito, J., Bath, T. S., et al. (2012). Isolation and proteomic characterization of the *Arabidopsis* Golgi defines functional and novel targets involved in plant cell wall biosynthesis. *Plant Physiol.* 159, 12–26. doi:10.1104/pp.111.193151
- Parsons, H. T., González Fernández-Niño, S. M., and Heazlewood, J. L. (2013). “Separation of the plant Golgi apparatus and endoplasmic reticulum by free-flow electrophoresis,” in *Plant Proteomics: Methods and Protocols, Second Edition*, eds J. V. Jorrin Novo, S. Komatsu, W. Weckwerth, and S. Weinkoop (New York: Humana Press), 527–539.
- Percy, A. J., Chambers, A. G., Yang, J., Domanski, D., and Borchers, C. H. (2012). Comparison of standard- and nano-flow liquid chromatography platforms for

- MRM-based quantitation of putative plasma biomarker proteins. *Anal. Bioanal. Chem.* 404, 1089–1101. doi:10.1007/s00216-012-6010-y
- Perkins, D. N., Pappin, D. J., Creasy, D. M., and Cottrell, J. S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20, 3551–3567. doi:10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2
- Persil-Cetinkol, Ö, Smith-Moritz, A. M., Cheng, G., Lao, J., George, A., Hong, K., et al. (2012). Structural and chemical characterization of hardwood from tree species with applications as bioenergy feedstocks. *PLoS One* 7:e52820. doi:10.1371/journal.pone.0052820
- Shen, H., Poovaiah, C. R., Ziebell, A., Tschaplinski, T. J., Pattathil, S., Gjersing, E., et al. (2013). Enhanced characteristics of genetically modified switchgrass (*Panicum virgatum* L.) for high biofuel production. *Biotechnol. Biofuels* 6, 71. doi:10.1186/1754-6834-6-71
- Swartz, M. E. (2005). UPLCTM: an introduction and review. *J. Liq. Chromatogr. Relat. Technol.* 28, 1253–1263. doi:10.1081/JLC-200053046
- Tyo, K. E., Alper, H. S., and Stephanopoulos, G. N. (2007). Expanding the metabolic engineering toolbox: more options to engineer cells. *Trends Biotechnol.* 25, 132–137. doi:10.1016/j.tibtech.2007.01.003

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 11 February 2015; accepted: 18 March 2015; published online: 01 April 2015.

Citation: González Fernández-Niño SM, Smith-Moritz AM, Chan LJG, Adams PD, Heazlewood JL and Petzold CJ (2015) Standard flow liquid chromatography for shotgun proteomics in bioenergy research. *Front. Bioeng. Biotechnol.* 3:44. doi: 10.3389/fbioe.2015.00044

This article was submitted to *Bioenergy and Biofuels*, a section of the journal *Frontiers in Bioengineering and Biotechnology*.

Copyright © 2015 González Fernández-Niño, Smith-Moritz, Chan, Adams, Heazlewood and Petzold. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.